Slow nucleic acid unzipping kinetics from sequence-defined barriers

S. Cocco[1], J.F. Marko[2], R. Monasson[3]

[1] CNRS–Laboratoire de Dynamique des Fluides Complexes, 3 rue de l'Université, 67000 Strasbourg, France;

[2] Department of Physics, The University of Illinois at Chicago, 845 West Taylor Street, Chicago, IL 60607-7059;

[3] CNRS–Laboratoire de Physique Théorique de l'ENS, 24 rue Lhomond, 75005 Paris, France.

Recent experiments on unzipping of RNA helix-loop structures by force have shown that $\approx$ 40-base molecules can undergo kinetic transitions between two well-defined 'open' and 'closed' states, on a timescale $\approx$ 1 sec [Liphardt *et al.*, Science 297, 733-737 (2001)]. Using a simple dynamical model, we show that these phenomena result from the slow kinetics of crossing large free energy barriers which separate the open and closed conformations. The dependence of barriers on sequence along the helix, and on the size of the loop(s) is analyzed. Some DNAs and RNAs sequences that could show dynamics on different time scales, or three(or more)-state unzipping, are proposed.

PACS: 87.15.-v, Biomolecules: structure and physical properties

I. INTRODUCTION

Helix-loops are the basic secondary-structure elements of folded single-stranded nucleic acids (ssNA). Recent physical studies of single helix-loop RNAs have revealed that despite their simple structures, they can display interesting dynamics [1]. When placed under moderate tensions $\approx$ 15 pN, telegraph-noise-like 'switching' behavior can be observed. The characteristic time of this switching has been seen to be on the $\approx$ 1 sec timescale, surprisingly large given the small size ($\approx$ 10 nm) of the molecules. The purpose of this paper is to present a simple theory capable of reproducing these slow switching kinetics.

The model we use for the energy of a helix-loop structure considers states of partial 'unzipping' [2]. It is based on available quantitative descriptions of base-pairing interactions and single-strand nucleic acid elastic response. In general a large free energy barrier between open and closed states for helix-loop structures is found, which qualitatively explains the observed two-state switching kinetics. We then combine this model with Eyring-Kramers transition-state theory to access the kinetics, and show how this barrier requires a long timescale to be crossed. Finally, we discuss how essentially the same approach can be used to describe branched-helix structures, and we predict three-state-switching for a specific molecular architecture.

II. MODEL OF BASE-PAIRING AND SSNA ELASTICITY

Our description uses the series of molecule configurations obtained by successively breaking base-pairs, starting from the molecule ends [2]. In addition to the base-pairing free energy of the double-stranded part of the molecule, we take into account the elastic response of the extended, unpaired, single-stranded part of the molecule [3–6]. The free energy change associated with opening base pair i is simply the difference between the free energies of the paired region, and the free energy associated with extension of the open region,

$$\Delta g(i,f) = g_0(i) - 2 g_s(f,i) \tag{1}$$

Here $g_0(i)$ is the free energy of opening base pair i (between 1 k_BT and 5 k_BT), obtained using the MFOLD program [7] . The stretching contribution, $g_s(f,i)$, is the free energy per base at constant force for stretched ssDNA, which provides a reasonable estimate for ssRNA. Integration of an empirical fit to the experimental ssDNA force-extension [8] gives: $g_s^{FJCL}(f) = k_B T l_{ss}/d \, \ln[\sinh(u)/u]$ with $u \equiv d\,f/[k_BT]$ and parameters $l_{ss} \simeq 5.6$ Å, $d = 15$ Å. Internal unpaired bases (Fig. 1) are considered to open along with the base pair immediately preceding them; those base opening steps therefore pick up a multiple-base g_s contribution. The free energy to unpair the first n base pairs is just the sum of the free energies of the individual base-opening steps,

$$G(n,f) = \sum_{i=1}^{n} \Delta g(i,f) \tag{2}$$

We emphasize that both the base-pairing and the elasticity contributions to (1) and (2) are free energies, i.e. are coarse-grained over atomic-scale fluctuations. The above zipper model [2] is roughly

1

equivalent to that described in the Supplementary Materials of [1]; the presence of the stretching force allows to discard complex opening/closing pathways relevant at zero force [9]. Note that some cooperativity between base pairs is introduced by MFOLD (the pairing free energy of neighboring base pairs is not simply additive, but includes some sequence specific stacking interactions).

III. EQUILIBRIUM BEHAVIOR OF A SINGLE HELIX-LOOP STRUCTURE

Fig. 2a shows $G(n, f)$ and the equilibrium probability $P(n, f) \propto \exp(-G(n, f)/k_B T)$ for a simple RNA hairpin, called P5ab [10] (Fig. 1a), under the solution conditions of [1] near the critical force $f^* \approx 15$ pN where two-state switching is observed. At the critical force, the open and closed states ($n = 22$ and $n = 3$, note n is just the number of broken base pairs) dominate; below or above this force, the free energy landscape is tilted either to favor the $n = 3$ or $n = 22$ state. At the critical force, there are various barriers due to drops in free energy resulting from openings of the U bulge, the weak non Watson-Crick GA central pairs, and the final loop. The largest barrier $\simeq 11$ k_BT must be crossed to reach the closed state from the open one, and vice-versa. The two-state behavior observed for P5ab follows from the partition of probability into two peaks separated by an improbably-accessed barrier region (see Supplementary Materials, Ref. [1]).

IV. DYNAMICAL MODEL

To reach a quantitative understanding of switching, we introduce a dynamical model for the motion of the 'fork' separating the base-paired and opened regions of the molecule [11]. We propose the following expressions for the elementary rates of opening and closing base pair n (*i.e.* to move the boundary between the open and closed portion of the molecule from n to $n - 1$ or to $n + 1$):

$$r_o(n) = r \, e^{-g_0(n)/k_B T} \quad , \qquad r_c(f, n) = r \, e^{-2g_s(f, n)/k_B T} \tag{3}$$

Here r is essentially the microscopic rate for a base pair to move together or apart in the absence of tension or base-pairing interactions, or roughly the inverse self-diffusion time for a few-nm-diameter object [12], $r\eta\ell^3/k_B T \approx 10^7 \text{ s}^{-1}$, with $\ell = 10$ nm, $\eta = 0.001$ Pa sec, and $k_B T = 4 \times 10^{-21}$ J.

In (3) we have made the simplifying approximation that the opening rate r_o has no force dependence, and is simply proportional to the exponential of the base-pairing free energy of (1). Eyring–Kramers transition–state theory applied to breaking of a chemical bond considers indeed the potential energy to be 'tilted' by a force-times-displacement contribution. Because hydrogen bonds break for relatively small displacements (≈ 0.1 nm) the reduction in the potential energy of the single-base-opening transition state will be roughly 15 pN $\times 0.1$ nm = 0.3 $k_B T$. This can be neglected with respect to the base-pairing free energy of a few $k_B T$, which is a lower bound to the energy of the transition state associated with breaking a single base pair. Detailed balance then determines the closing rate r_c to be proportional to the exponential of force times displacement, *i.e.* to the energy of a fluctuation that is able to pull the two bases back together in opposition to the applied force.

The rates (3) lead to a master equation for the probability $\rho_n(t)$ for the boundary to be at site n at time t:

$$\frac{d\rho_n(t)}{dt} = -\sum_{m=0}^{N} T_{n,m} \, \rho_m(t) \tag{4}$$

This $(N+1) \times (N+1)$ matrix $T_{n,m}$ is tridiagonal, with nonzero entries $T_{m-1,m} = r_c(f, m)$, $T_{m+1,m} = r_o(m)$, and $T_{m,m} = -T_{m-1,m} - T_{m+1,m}$.

V. SWITCHING KINETICS OF A SINGLE HELIX-LOOP STRUCTURE

We have solved (diagonalized) (4) for P5ab (Fig. 1a). The smallest eigenvalue is 0; the eigenvector is the equilibrium Boltzmann distribution. At the critical force where the molecule is on average half-open, the smallest non zero eigenvalue is $\lambda_1 = 2.1 \times 10^{-6} r$, which corresponds to the slowest mode of fluctuation, the 'switching' of the boundary of the open region between $n \approx 3$ and $n \approx 22$. The remaining 21 eigenvalues are all well separated from the leading ones. The second largest

2

eigenvalue is $\lambda_2 = 0.9 \times 10^{-4} r$ (Table). Thus the theoretical dynamics of P5ab involve one slow opening-closing transition, combined with many other transitions occurring more than 50 times faster. The net rates of opening (k_o) and closing (k_c) can be computed from $\lambda_1 = (k_o + k_c)$, and the ratio of the open and closed equilibrium probabilities equal to k_c/k_o. To compare our theoretical results with the experiments of [1] we have fitted our result for $k^* \equiv k_o = k_c$ at the force f^* where the open and closed states have equal probability to experimental data [13], giving $r = 3.6 \times 10^6$ sec^{-1}.

Fig. 3a shows time series from Monte Carlo simulations of the molecular motion; slow two-state switching is seen on a ≈ 0.25 sec timescale, on top of which occur much faster small fluctuations. When we convolve these data with a 20 Hz low-pass filter (as used experimentally [1]) the result (Fig. 3b) is essentially the same as the experiment. The variation of the rates with force given by theory are also in good agreement with experiment(Fig. 4). The 'transition state' coordinate can be inferred from the relative slopes of the opening and closing rates of Fig. 4 (independent of the fitted parameter r) around the critical force [1]. The critical-force transition state is at the top of the free energy barrier, at $n^* = 11 - 12$ (Fig. 2a).

VI. BARRIERS FROM GENERIC HELIX-LOOP STRUCTURE

For molecules with repeated sequences and no terminal loop e.g. AU followed by a long GC helix or a crosslink between end bases, the kinetics essentially corresponds to a diffusion in a flat free energy landscape, and shows none of the two-state character of the P5ab RNA. For a repeated AU sequence of $n = 25$ base pairs, the switching time is just the diffusion-like time $t^* \approx 2N^2/(\pi^2 r_o^*) = 2 \times 10^{-4}$ sec (Table). Thus, the 1000-times longer switching time of P5ab comes from the 11 k_BT barrier of Fig. 2a.

The presence of a loop is sufficient to generate such a large barrier and, consequently, two-state switching. The simplest illustration is given by a homogenous DNA sequence ending with a loop. We have considered a 24-base Poly(GC) homogeneous-sequence helix terminated with a 4-base loop (Fig. 1b, 2b) [14]. Our theory predicts a two-state switching behavior on the same time scale as P5ab. The switching time is 50 times larger for a longer 8-base loop (Table).

The free energy barrier G^* at criticality for a S-base-pair stem (uniform pairing free energy g_0) followed by a L-base loop (closing free energy $g_{loop}(L)$ at zero force) (Fig. 1) can be simply estimated. The critical force f^* is given by the condition that the free energy of the open molecule equals the free energy of the closed molecule, $G(0, f^*) = G(S, f^*) = S\, g_0 - (2S + L)\, g_{ss}(f^*) - g_{loop}(L)$. The barrier height $G^*(S, L) \equiv G(S - 1, f^*)$ then reads

$$G^*(S, L) = (S - 1)(g_0 - 2\, g_{ss}(f^*)) = \frac{(S - 1)\, (g_0\, L + 2\, g_{loop}(L))}{L + 2S} \quad . \tag{5}$$

Table 1 shows that, for a fixed stem length S, the critical force decreases with the length L of the loop, while the free energy barrier G^*, and the switching time t^* increase. For non-repeated sequences, (5) is only approximate when substituting g_0 with an average pairing free energy; it allows an estimate of how the switching time depends on S and L [15]. Note that the critical barrier essentially depends on the smaller of the two lengths S, L.

VII. BRANCHED-HELIX MOLECULES AND MULTIPLE-STATE-SWITCHING

Our approach can be extended to more complicated situations e.g. nucleic acids with branched structure. An example is the P5abcΔA RNA molecule [1] of Fig. 1c, with free energy landscape as shown in Fig. 2c. The opening of the first 12 bases (helix H1) follows as above, but going past the branch, description of the independent opening of the two helix regions requires a three-dimensional free energy landscape *i.e.* free energy as a function of the positions of the two unzipping boundaries [16]. A rich behavior emerges, shown in Fig. 3c. At the critical force $f^* = 12.9$ pN, H1 switches on a long time scale $t^* = 10$ sec (Table), while the short lateral helix (H2) opens and closes with a much shorter characteristic time $t_2 \simeq 9$ msec. The lateral long helix (H3) opens very rarely at this force. Predictions for the rates, barrier height, ... are reported in Table 1.

It is possible to design molecules with multiple-state dynamics. Consider the molecule of Fig. 1d, which has two well-bound Poly(GC) regions separated by an unpaired bubble, and terminated by a loop. For this system, three-state switching occurs (Fig. 3d), and should be observable in the frequency range accessible in experiments such as that of [1].

3

VIII. CONCLUSION

A few improvements might be added to this model. First, opening may occur through the cooperative nucleation of a few base pairs bubble [4]. Second, mismatches might take place during closing [17], although they are highly limited by the presence of the 15 pN force. Finally it would be interesting to be able to describe unzipping events involving breaking of tertiary structures; these are thought to be present in the molecule P5abcΔA in presence of Mg^{2+} [1,18].

A general result of our work is that slow switching character should be quite generic for small biological molecules with helix loop structure. We note that our approach could also be used to analyze the opening-closing dynamics of nucleic-acid-detecting DNA 'beacons' [19], both on their own, and in the presence of their targets. In this case the 'unzipping' forces are applied by the hybridization interactions instead of by a large force transducer. Since such experiments amount to molecule recognition processes it is not impossible that slow barrier-crossing transitions of the sort discussed here occur in-vivo.

IX. ACKNOWLEDGEMENTS

Molecule (N)	ΔG k_BT	f^* (pN)	$\ln(k^*$ /r)	t^* (sec)	$\ln k_o(f)$ (sec^{-1})	$\ln k_c(f)$ (sec^{-1})	G^* (k$_B$T)	λ_1/λ_2
P5ab (49)	57.2 **66.7** **±8.5**	15.1 **14.5** **±1**	-13.8	0.25 **0.25** **±0.1**	-42.9+1.93 f **-39±2.3 −** **(2.9±0.2)f**	27.5 -2.74 f **41±1.9 +** **(2.8±0.1)f**	10.7	0.023
Repeated AU (50) GC (50)	44.8 135	12.3 27.4	-6.7 -10.3	0.0002 0.008			0 0	0.99 0.99
Poly(GC) no loop (24) 4b loop (28)	41.3 34.8 **43±3**	19.7 15.6 **16**	-6.5 -14.1	0.0002 0.37	-46.7 + 2.02 f	1.7 - 1.01 f	2.4 11.3	0.42 2 10^{-4}
8b loop (32) double (56)	33.2 71.3	13.7 15.8	-18.0 -14.4	16.7 0.48	-46.8 + 2.1 f -53.2+2.26 f	1.3 -1.4 f 8.34-1.44 f	15.7 10.9	3 10^{-6} 0.38 ^{a}
P5abcΔA (64)	70.6 **71.9** **±11.5**	12.9 **12.7** **±0.3**	-17.1	10^b	-43.8 + 2.06 f **-39±9.3 +** **(2.7±0.7)f**	9.4 -2.05 f **58±7.5 −** **(4.2±0.5)f**	14.3	9 10^{-4}

TABLE I. Theoretical results for the molecules of Fig. 1 compared to experimental values from [1, 14] (in bold) when available. Columns indicate the free energy $\Delta G = G(N, f = 0)$ at zero force, the critical force f^*, the rate of opening/closing k^*, the switching time t^* and the free energy G^* of the highest barrier at criticality, the variation of the opening and closing rates upon force, and the ratio of the two largest non zero eigenvalues. This ratio is small when open and closed states are well defined, and close to one otherwise. Uncertainties in base-pairing free energy are at most $\delta g \approx 0.5 k_B T/$bp, with consequent total uncertainties of $\approx N^{1/2}\delta g$ (or about $3 k_B T$ for $N = 25$) for ΔG and of $\approx (N/2)^{1/2}\delta g \simeq 2 k_B T$ for G^*, and $\approx 3 k_B T/20$ nm≈ 0.6 pN for the critical force. Notes: ^{a} the ratio is close to one due to the presence of three, and not two, states giving rise to two large barrier crossing times (the remaining fluctuation times are much shorter: $\lambda_1/\lambda_3 \simeq 0.0001$); ^{b} the predicted rate (0.1/sec) for P5abcΔA is very close to the lowest frequency (0.05 Hz) resolved experimentally [1], thus its value is known only roughly.
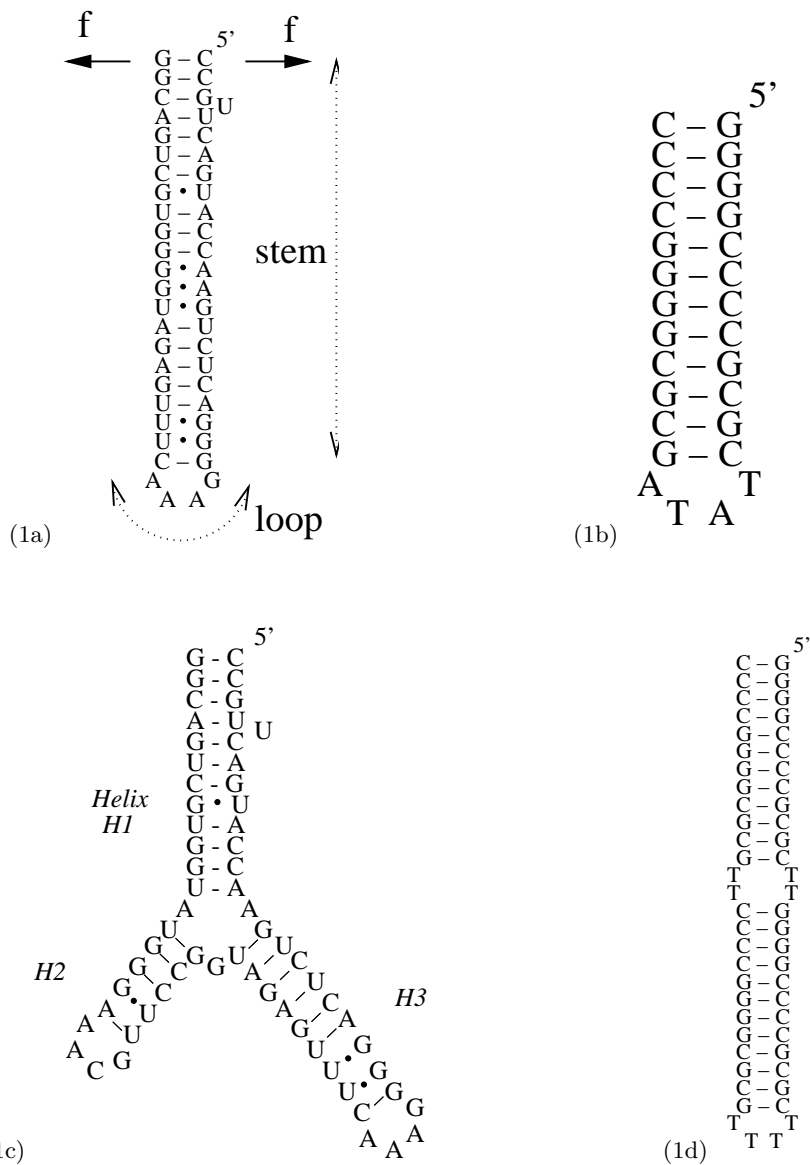
(1a)

f ← f →

5'

G – C
G – C
C – G U
A – U
U – C
U – A
C – G
G • U
U – A
G – C
G – C
G • A
G • A
U • G
A – U
G – C
A – U
G – C
U – A
U • G
U • G
C – G

A   G
  A   A

stem

loop

(1b)

5'

C – G
C – G
C – G
C – G
C – G
G – C
G – C
G – C
G – C
C – G
G – C
C – G
G – C

A   T
  T   A

(1c)

5'

G - C
G - C
C - G
A - U
G - C U
U - A
C - G
G • U
U - A
G - C
G - C
U - A

Helix H1

A   A

H2   *H3*

(1d)

5'

C – G
C – G
C – G
C – G
G – C
G – C
G – C
G – C
C – G
C – G
C – G
G – C

T   T
T   T

C – G
C – G
C – G
C – G
G – C
G – C
G – C
G – C
C – G
G – C
G – C
C – G

T   T   T

FIG. 1. Nucleic acid unzipping experiment and molecules studied. A constant force is applied to a helix-loop structure while the distance between molecule ends is measured. (a) P5ab RNA, a single helix-loop structure present on the P4-P6 domain of a self splicing group I intron of the *Tetrahymena thermophila*. The structure shown is predicted by Mfold [7] apart from the G-A weak pairs [10] indicated here with dots, and with the U-bulge translocated; (b) DNA hairpin consisting of a poly(GC) helix terminated with an ATAT loop [14]; (c) P5abcΔA RNA, a variant of P5ab with an additional helix giving a Y-branched structure at zero force; (d) Hypothetical RNA molecule obtained by ligation of two poly(GC) helices as in (b) and replacing A bases with Ts.
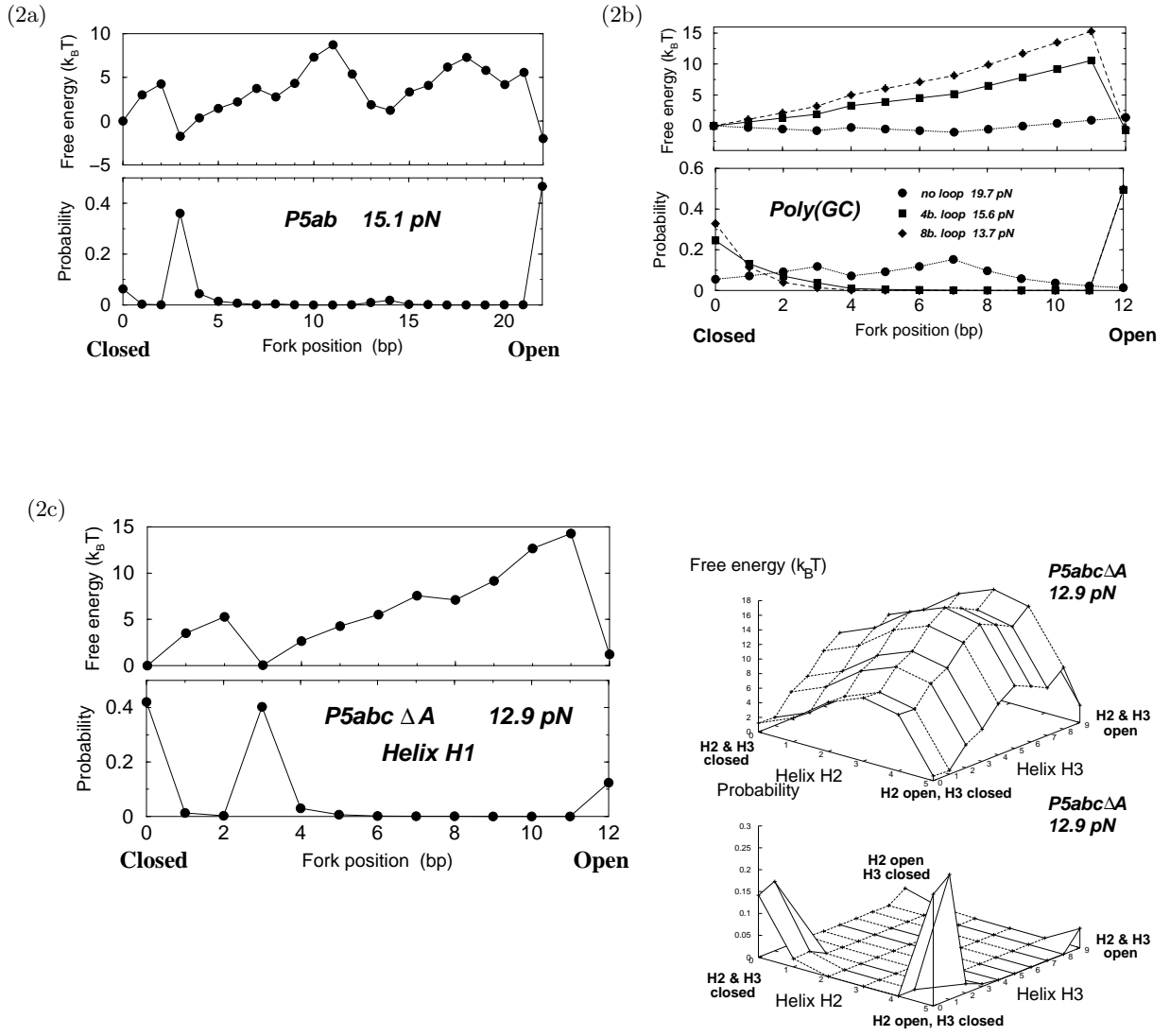
FIG. 2. Free energy landscapes (in k_BT, top) and probability distributions for the position of the opening fork (bottom) at the critical force for molecules of Fig. 1. (a) P5ab; (b) Poly(GC) RNA with: a 8bp loop (diamond), a 4 bp loop (squares); no loop (circles). (c) P5abcΔA; (left: opening positions $n_1 = 0, \ldots, 12$ corresponding to helix H1, right: opening positions (n_2, n_3) with $n_2 = 0, 5$ – H2 – and $n_3 = 0, 9$ – H3 –); by definition configurations $n_1 = 12$ and $(n_2 = 0, n_3 = 0)$ coincide.
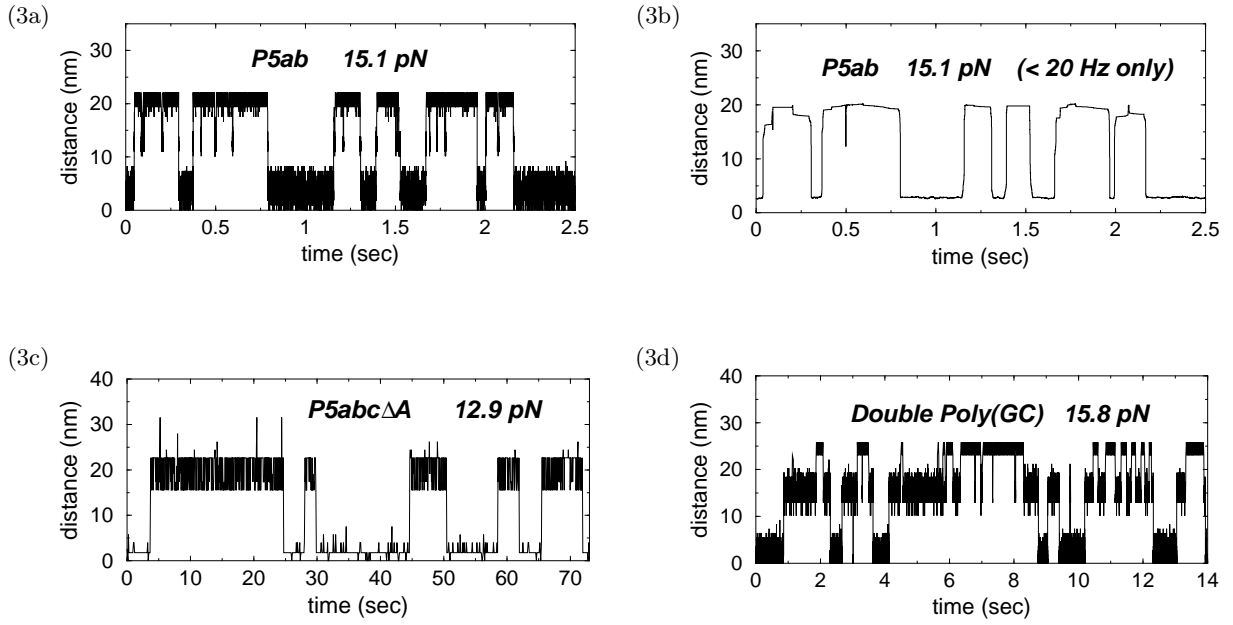
FIG. 3. Unzipping kinetics at the critical force. Distances between extremities of the molecule are shown for: (a) P5ab; there is a slow switching between the $n \simeq 3$ and $n \simeq 22$ configurations and fast transitions between configurations n around these ones (Fig 2a). (b) Convolution of (a) where oscillations faster than 20Hz are averaged out; (c) P5abcΔA; there is a slow switching between the closed molecule and the molecule with $H1$ opened and $H2$ opening or closing on a shorter time scale, the opening of $H3$ (distance between extremities of 31 nm) is a rare event at the force of 12.9 pN; (d) the hypothetical RNA molecule of Fig. 1d.
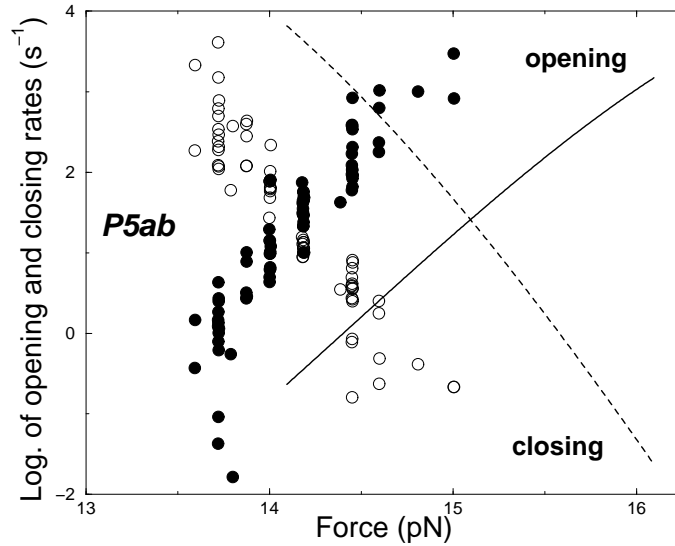

FIG. 4. Log. of the opening and closing rates for P5ab as measured in [1] as a function of force (full circles: opening, empty circles: closing), compared to theory (full line: opening, dashed line: closing). The slopes of $\ln k_o, \ln k_c$ give the relative positions $n_o = 8, n_c = 11$ of the transition state from the closed ($n = 3$) and open ($n = 24$) states respectively, with an absolute location in $n^* \simeq 12$.

7

[1] Liphardt, J., Onoa, B., Smith, S.B., Tinoco, I.Jr. *Science* **297**, 733 (2001).

[2] Poland, D., & Scheraga, H.A. *Theory of Helix-Coil Transitions*, Academic Press (1970).

[3] Bockelmann, U., Essevaz-Roulet, B.& Heslot, F. *Phys. Rev. E* **58**, 2386 (1997).

[4] Cocco, S., Monasson, R. & Marko, J.F. *Proc. Natl. Acad. Sci. USA* **98**, 8608 (2001); *Phys. Rev. E* **65**, 041907 (2002).

[5] Gerland, U., Bundschuh, R. & Hwa, T. *Biophys. J.* **81**, 1324-32 (2001).

[6] Lubensky, D.K. & Nelson, D.R. *Phys. Rev. Lett.* **85**, 1572 (2000); *Phys. Rev. E* **65**, 031917 (2002).

[7] Zuker, M. *Curr. Opin. Struct. Biol* **10**, 303 (2000).
To reproduce the experimental ionic conditions (10 mM Tris, 250 mM Na+, 10mM Mg2+ or EDTA), a correction on the free energy equal to - 0.193 $k_B T \ln([Na^+] + 3.3[Mg]^{0.5})$ (suggested by M. Zuker) has been applied. Pairing free energies for Non Watson-Crick GA base pairs have been obtained from their GU counterparts, with $\simeq 1$ k_BT corrections measured in Meroueh, M. & Chow, C.S. *Nucl. Acids Res.* **27**, 1118-25 (1999). Calculations for DNA sequence have been performed on DNA MFOLD server, at room temperature and 0.15 Na^+, using free energy parameters from Santa Lucia, J.Jr. *Proc. Nat. Aca. Sci. USA* **95**, 1460-1465 (1998).

[8] Smith, S.B., Cui,Y., & Bustamante, C. *Science* **271**, 795 (1996).

[9] Isambert, H. & Siggia, E.D. *Proc. Natl. Acad. Sci. USA*, **97**, 6515 (2000).
Zhang, W. & Chen, S.J. *Proc. Natl. Acad. Sci. USA*, **99**, 1931 (2002).

[10] Cate, J.H. *et al. Science* **273**, 1678 (1996).

[11] Kramers, H.A. *Physica* (Utrecht) 7, 284 (1940).

[12] Doi, M., & Edwards, S.F. *The Theory of Polymer Dynamics*. Clarendon, Oxford (1986).

[13] Possible effects of the feedback mechanism on the force used in [1] (private communication) are not considered here, but could be included in our model.

[14] Bustamante, C. et al. *Curr. Opin. Struct. Biol.* **10**, 279-85 (2000).

[15] Additional contributions to G^* are present for long, random sequences [6]. For uncorrelated base pair distributions of the free energies $g_0(i)$, the free energy $G(n, f)$ (2) behaves as a random walk, resulting in a maximal barrier $G^* = O(\sqrt{N})$.

[16] P5abcΔA includes three helices with $N_1 = 12$ (H1), $N_2 = 5$ (H2), and $N_3 = 9$ (H3) base pairs. A single (respectively double) boundary between open and closed regions arises if less (resp. more) than 12 base pairs are open. The number of configurations of the molecule is thus $N = N_1 + (N_2+1)(N_3+1) = 72$. From configurations where H2 and H3 are partially opened, four transition steps are possible.

[17] Ansari, A., Kuznetsov, S.V., & Shen, Y. *Proc. Natl. Acad. Sci. USA* **98**, 7771 (2001).

[18] Fang, X., Pan, T., & Sosnick, T.R. *Biochemistry*, **38(51)**, 16840 (1999).

[19] Bonnet, G., Tyagi, S., Libchaber, A. & Kramer F.R. *Proc. Natl. Acad. Sci. USA* **96**, 6171 (1999).